



## Measure Information

This document contains the information submitted by measure developers/stewards, but is organized according to NQF's measure evaluation criteria and process. The item numbers refer to those in the submission form but may be in a slightly different order here. In general, the item numbers also reference the related criteria (e.g., item 1b.1 relates to sub criterion 1b).

### Brief Measure Information

**NQF #:** 3751

**Corresponding Measures:**

**Measure Title:** Risk Adjusted Post-Ambulance Provider Triage Emergency Department (ED) Visit Rate Measure

**Measure Steward:** Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (CORE)

**sp.02. Brief Description of Measure:**

The Risk Adjusted Post-Ambulance Provider Triage Emergency Department (ED) Visit Rate Measure (shorthand: Post-Triage ED Visit Rate Measure) assesses the quality of the triage and decision making by ambulance providers who transport low acuity patients to an alternative destination (non-ED location), or facilitate Treatment In Place (TIP), by identifying whether patients have a subsequent ED visit or death within three days. The cohort includes adult Medicare patients. This measure is novel in that it is the only existing risk-adjusted outcome measure developed for quality measurement of ambulance providers and suppliers.

This measure is intended for use in the Emergency Triage, Treat, and Transport (ET3) Model from the Centers for Medicare and Medicaid Innovation. ET3 is a voluntary payment model that aims to improve quality and lower costs by reducing avoidable transports to the ED and unnecessary hospitalizations following those transports. Under this model, CMS will pay ambulance providers to 1) Transport to an Alternative Destination (TAD), such as a primary care office, urgent care clinic, or a community mental health center (CMHC), or 2) initiate and facilitate TIP with a qualified health care partner (QHCP), either at the scene of the 911 emergency response or via telehealth. CMS will continue to pay to transport a Medicare FFS beneficiary to a hospital emergency department or other covered destination. The Post-Triage ED Visit Rate Measure will capture the quality of the triage decision by measuring the number of ED visits or deaths within three days among patients who received TAD or TIP Interventions. The measure Flow Chart of the ET3 process begins with 911 call/initiation of ambulance service, then highlights the measure cohort (beneficiaries receiving TAD/TIP), then describes the measure numerator/outcome (beneficiaries who died/had ED visit within 3 days). The final box is a green check mark for success (the beneficiary did not die/no ED visit within 3 days). The measure is designed to promote high quality care by ambulance service providers by ensuring that, as ET3 Model service use increases, triage decisions still ensure the safe delivery of care regardless of care setting.

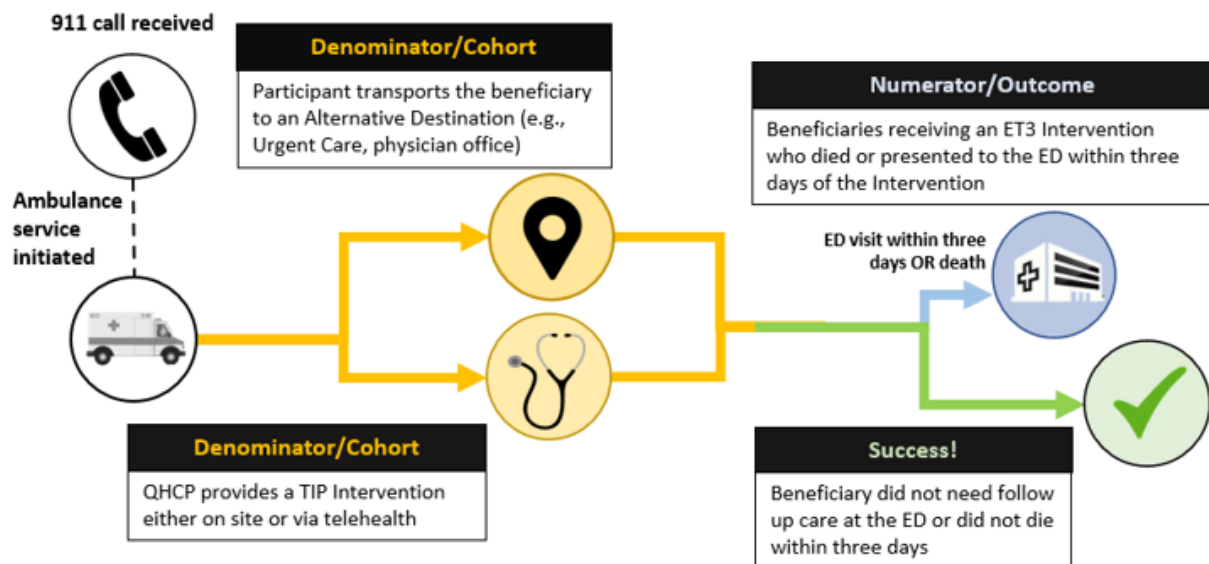


Figure 1 ET3 Measure Flow Chart

#### 1b.01. Developer Rationale:

**sp.12. Numerator Statement:** This is a risk-adjusted outcome measure. The outcome for this measure is an ED visit or death within three days for patients who have been triaged by an ambulance provider to an alternative non-ED destination or treated in place (TAD/TIP). ED visits include observation stays or hospital admissions first evaluated through the ED. Patients directly admitted to hospital inpatient or observation care without receiving ED services are not counted as outcome events. Patients who visit the ED within three days but are discharged with a primary diagnosis related to mental health or substance-use disorder are not counted as outcome events.

**sp.14. Denominator Statement:** The cohort, or denominator, includes patients age 18 or older who have an encounter with an ambulance provider whose triage decision is to either transport them to an alternative non-ED destination (i.e., TAD) or to initiate and facilitate TIP. The measure does not include patients who are enrolled in hospice care at the time of the TAD/TIP encounter with the ambulance provider.

**sp.16. Denominator Exclusions:** The measure has no exclusions.

**Measure Type:** Outcome

**sp.28. Data Source:**

Claims

Other (specify)

Medicare Enrollment Data and other administrative data.

**sp.07. Level of Analysis:**

Other

**IF Endorsement Maintenance – Original Endorsement Date:**

**Most Recent Endorsement Date:**

**IF this measure is included in a composite, NQF Composite#/title:**

**IF this measure is paired/grouped, NQF#/title:**

**sp.03. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?:**

## 1. Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria

---

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Measure and Report: Evidence section. For example:

**Current Submission:**

Updated evidence information here.

**Previous (Year) Submission:**

Evidence from the previous submission here.

**1a.01. Provide a logic model.**

*Briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.*

[Response Begins]

[Response Ends]

**1a.02. Provide evidence that the target population values the measured outcome, process, or structure and finds it meaningful.**

*Describe how and from whom input was obtained.*

[Response Begins]

[Response Ends]

**1a.03. Provide empirical data demonstrating the relationship between the outcome (or PRO) and at least one healthcare structure, process, intervention, or service.**

[Response Begins]

[Response Ends]

**1b.01. Briefly explain the rationale for this measure.**

*Explain how the measure will improve the quality of care, and list the benefits or improvements in quality envisioned by use of this measure.*

[Response Begins]

[Response Ends]

**1b.02. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.**

*Include mean, std dev, min, max, interquartile range, and scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*

[Response Begins]

[Response Ends]

**1b.03. If no or limited performance data on the measure as specified is reported above, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement. Include citations.**

[Response Begins]

[Response Ends]

**1b.04. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.**

*Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included. Include mean, std dev, min, max, interquartile range, and scores by decile. For measures that show high levels of performance, i.e., “topped out”, disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*

[Response Begins]

[Response Ends]

**1b.05. If no or limited data on disparities from the measure as specified is reported above, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in above.**

[Response Begins]

[Response Ends]

## 2. Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.

---

**sp.01. Provide the measure title.**

*Measure titles should be concise yet convey who and what is being measured (see [What Good Looks Like](#)).*

**[Response Begins]**

Risk Adjusted Post-Ambulance Provider Triage Emergency Department (ED) Visit Rate Measure

**[Response Ends]**

**sp.02. Provide a brief description of the measure.**

*Including type of score, measure focus, target population, timeframe, (e.g., Percentage of adult patients aged 18-75 years receiving one or more HbA1c tests per year).*

**[Response Begins]**

The Risk Adjusted Post-Ambulance Provider Triage Emergency Department (ED) Visit Rate Measure (shorthand: Post-Triage ED Visit Rate Measure) assesses the quality of the triage and decision making by ambulance providers who transport low acuity patients to an alternative destination (non-ED location), or facilitate Treatment In Place (TIP), by identifying whether patients have a subsequent ED visit or death within three days. The cohort includes adult Medicare patients. This measure is novel in that it is the only existing risk-adjusted outcome measure developed for quality measurement of ambulance providers and suppliers.

This measure is intended for use in the Emergency Triage, Treat, and Transport (ET3) Model from the Centers for Medicare and Medicaid Innovation. ET3 is a voluntary payment model that aims to improve quality and lower costs by reducing avoidable transports to the ED and unnecessary hospitalizations following those transports. Under this model, CMS will pay ambulance providers to 1) Transport to an Alternative Destination (TAD), such as a primary care office, urgent care clinic, or a community mental health center (CMHC), or 2) initiate and facilitate TIP with a qualified health care partner (QHCP), either at the scene of the 911 emergency response or via telehealth. CMS will continue to pay to transport a Medicare FFS beneficiary to a hospital emergency department or other covered destination. The Post-Triage ED Visit Rate Measure will capture the quality of the triage decision by measuring the number of ED visits or deaths within three days among patients who received TAD or TIP Interventions. The measure Flow Chart of the ET3 process begins with 911 call/initiation of ambulance service, then highlights the measure cohort (beneficiaries receiving TAD/TIP), then describes the measure numerator/outcome (beneficiaries who died/had ED visit within 3 days). The final box is a green check mark for success (the beneficiary did not die/no ED visit within 3 days). The measure is designed to promote high quality care by ambulance service providers by ensuring that, as ET3 Model service use increases, triage decisions still ensure the safe delivery of care regardless of care setting.

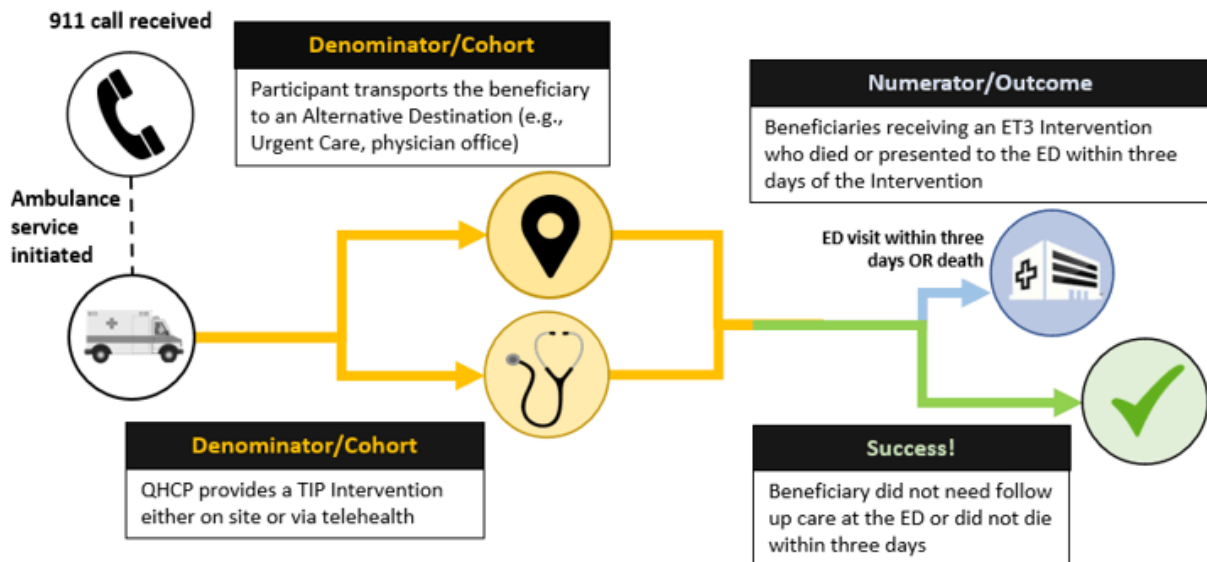


Figure 1 ET3 Measure Flow Chart

**[Response Ends]**

**sp.04. Check all the clinical condition/topic areas that apply to your measure, below.**

*Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.*

*Please do not select:*

- Surgery: General

**[Response Begins]**

Other (specify)

**[Other (specify) Please Explain]**

Acute Care

**[Response Ends]**

**sp.05. Check all the non-condition specific measure domain areas that apply to your measure, below.**

**[Response Begins]**

Care Coordination

Care Coordination: Transitions of Care

Safety

**[Response Ends]**

**sp.06. Select one or more target population categories.**

*Select only those target populations which can be stratified in the reporting of the measure's result.*

*Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.*

*Please do not select:*

- *Populations at Risk: Populations at Risk*

**[Response Begins]**

Adults (Age >= 18)

**[Response Ends]**

**sp.07. Select the levels of analysis that apply to your measure.**

*Check ONLY the levels of analysis for which the measure is SPECIFIED and TESTED.*

*Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.*

*Please do not select:*

- *Clinician: Clinician*
- *Population: Population*

**[Response Begins]**

Other

**[Response Ends]**

**sp.08. Indicate the care settings that apply to your measure.**

*Check ONLY the settings for which the measure is SPECIFIED and TESTED.*

**[Response Begins]**

Inpatient/Hospital

Other

**[Response Ends]**

**sp.09. Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials.**

*Do not enter a URL linking to a home page or to general information. If no URL is available, indicate "none available".*

**[Response Begins]**

None available.

**[Response Ends]**

**sp.12. Attach the data dictionary, code table, or value sets (and risk model codes and coefficients when applicable). Excel formats (.xlsx or .csv) are preferred.**

*Attach an excel or csv file; if this poses an issue, [contact staff](#). Provide descriptors for any codes. Use one file with multiple worksheets, if needed.*



**[Response Begins]**

Available in attached Excel or csv file

**[Response Ends]**

Attachment: 3751\_DataDictionary\_ET3\_NQF Submission.xlsx

For the question below: state the outcome being measured. Calculation of the risk-adjusted outcome should be described in sp.22.

**sp.13. State the numerator.**

*Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome).*

*DO NOT include the rationale for the measure.*

**[Response Begins]**

This is a risk-adjusted outcome measure. The outcome for this measure is an ED visit or death within three days for patients who have been triaged by an ambulance provider to an alternative non-ED destination or treated in place (TAD/TIP). ED visits include observation stays or hospital admissions first evaluated through the ED. Patients directly admitted to hospital inpatient or observation care without receiving ED services are not counted as outcome events. Patients who visit the ED within three days but are discharged with a primary diagnosis related to mental health or substance-use disorder are not counted as outcome events.

**[Response Ends]**

For the question below: describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in sp.22.

**sp.14. Provide details needed to calculate the numerator.**

*All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets.*

*Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.*

**[Response Begins]**

The outcome for this risk-adjusted measure is an ED visit death within three days among patients triaged by an ambulance provider to an alternative non-ED destination or treated in place (TAD/TIP). Patients seeking higher acuity emergency care within a few days of having been triaged to lower acuity care may be a sign of poor quality of triage care. To guard against extreme cases where the quality of care may have caused a patient's death by delaying necessary emergency care, the measure counts all-cause mortality as an outcome.

#3751 Risk Adjusted Post-Ambulance Provider Triage Emergency Department (ED) Visit Rate Measure ,  
Submission Last Updated: Jan 11, 2023

ED visits are identified by facility-based ED claims (revenue center codes: '0450','0451','0452','0456','0459','0981'). Mortality outcomes are identified using the Medicare Beneficiary Summary Files (MBSF) in Chronic Condition Warehouse (CCW), using the date of death from CMS Common Medicare Environment (CME).<sup>1</sup>

Outcome Attribution: Each TAD/TIP encounter is attributed to the ambulance provider who received a Medicare Fee-For-Service (FFS) payment for TAD/TIP. If multiple ambulance providers received payment for TAD/TIP encounters for the same patient on the same day, both providers are held accountable for the patient's outcome, with the patient in that case attributed to both ambulance providers.

**Numerator Exclusions:**

To ensure the validity of the measure, several events are not counted within the outcome as they represent events that can be feasibly captured and are more likely to be outside of the control of ambulance providers:

1. Patients who visit the ED within three days of TAD/TIP Intervention but are discharged with a primary diagnosis related to mental health or substance use disorder are excluded from the numerator.

**Rationale:** This approach aims to incentivize the use of intervention (TAD/TIP) among all patients when appropriate, including patients with mental health or substance-use disorders who tend to be higher users of ambulance and ED care, without penalizing the ambulance providers/suppliers who are providing TAD/TIP if a subsequent ED visit results from their mental health or substance use needs.

Mental health or substance-use disorders are identified using CMS Hierarchical Condition Category Codes (HCC) and one individual ICD-10 code: R45851 Suicidal Ideation; from HCC Minor Symptoms, Signs, Findings, modified. The full list of ICD-10 codes contained within these HCCs is in the accompanying data dictionary Excel file. [Table 1](#) below contains the HCC codes.

2. If patients receive multiple TAD/TIP Interventions within three days before an ED visit or death, the ED visit or death *only* counts as an outcome event for the proximal intervention, and the patient is attributed to the ambulance provider associated with that encounter.

**Rationale:** This approach avoids attributing the outcome to multiple TAD/TIP intervention encounters that are likely to be performed by the same ambulance provider, thereby avoiding double counting of an outcome.

Hierarchical Condition Category (HCC)	Condition Category Label
54	Substance Use with Psychotic Complications
55	Substance Use Disorder, Moderate/Severe, or Substance Use with Complications
56	Substance Use Disorder, Mild, Except Alcohol and Cannabis
57	Schizophrenia
58	Reactive and Unspecified Psychosis
59 <sup>1</sup>	Major Depressive, Bipolar, and Paranoid Disorders (Except 360 sequela codes; see Data Dictionary Excel)
60	Personality Disorders
61	Depression
62	Anxiety Disorders
63	Other Psychiatric Disorders
202	Drug Use, Uncomplicated, Except Cannabis
203	Alcohol/Cannabis Use or Use Disorder, Mild or Uncomplicated; Non-Psychoactive Substance Abuse; Nicotine Dependence

Table 1 Hierarchical Condition Category Codes (HCC) of Mental Health and Substance Use Disorder  
Not Counted in the Measure Outcome

<sup>1</sup>360 codes removed from CC59, which were 'sequela codes', defined as the residual effect (condition produced) after the acute phase of an illness or injury has terminated. These sequela codes are distinct from initial encounter codes that reflect an acute or initial healthcare need. These are unlikely to be coded as a primary diagnosis due to the nature of sequela codes, and clinically, these are rarely deemed to be the cause of an acute visit.

Reference:

1. CODEBOOK: Medicare Beneficiary Summary File (MBSF) Base with Medicare Part A, B, C, and D. Chronic Condition Warehouse. February 2021. Version 1.4.

**[Response Ends]**

For the question below: state the target population for the outcome. Calculation of the risk-adjusted outcome should be described in sp.22.

**sp.15. State the denominator.**

*Brief, narrative description of the target population being measured.*

**[Response Begins]**

The cohort, or denominator, includes patients age 18 or older who have an encounter with an ambulance provider whose triage decision is to either transport them to an alternative non-ED destination (i.e., TAD) or to initiate and facilitate TIP. The measure does not include patients who are enrolled in hospice care at the time of the TAD/TIP encounter with the ambulance provider.

**[Response Ends]**

For the question below: describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in sp.22.

**sp.16. Provide details needed to calculate the denominator.**

*All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets.*

*Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.*

**[Response Begins]**

Patients are eligible for inclusion in the measure if they are:

- Enrolled in Medicare FFS Part A and Part B for at least 6-months (out of the 12-months prior) prior to TAD/TIP encounter plus coverage at time of intervention and one-month post-intervention for outcome identification.

- **Rationale:** 6-month enrollment is required for claims-based risk-adjustment to adequately identify comorbidities and other risk variables.
- Aged 18 years and older.
  - **Rationale:** The ET3 Model includes adult patients.
- Have an encounter with an ambulance provider where patient was triaged to TAD/TIP. The patients can have multiple TAD/TIP encounters, and all encounters will be included in the cohort. However, if multiple EMS encounters occur on the same day, we choose one encounter.
  - **Rationale:** Supports measure intent and cohort is encounter based. For multiple same-day encounters, we include only one encounter because there is no timestamp information on EMS claims and we are unable to determine which event occurred first.
- Are not enrolled in Medicare hospice at the time of TAD/TIP encounter with the Model Participant. Hospice status is identified through enrollment in Medicare hospice services in the patient enrollment file.
  - **Rationale:** Patients in hospice care have complex medical needs and have an outcome rate unrelated to ambulance provider decision-making or quality of care. Excluding these patients ensures that model participants still offer TAD/TIP services to these beneficiaries, which are likely to be highly aligned with hospice service beneficiary preferences. Also addressing ambulance provider concerns of higher-than-expected ED utilization or death among these beneficiaries within 3 days.

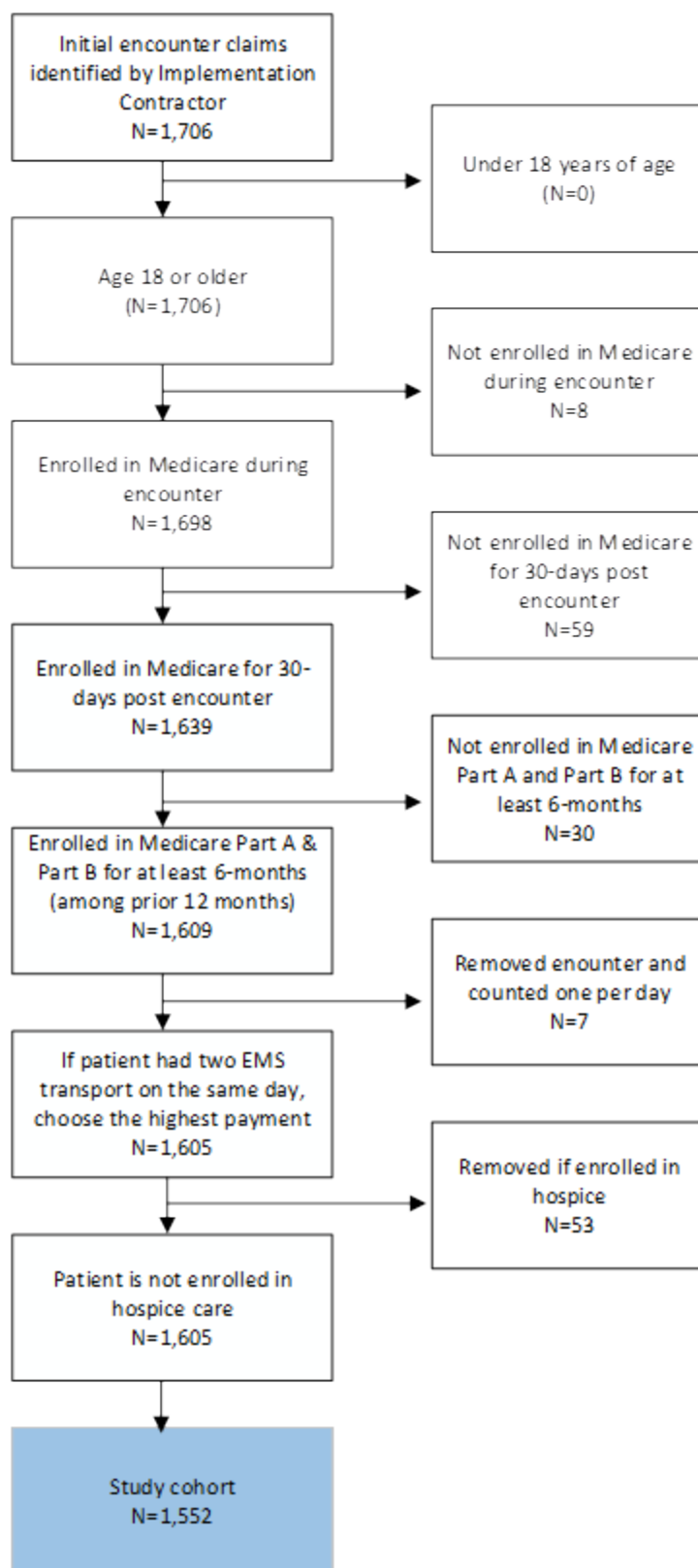


Figure 2 ET3 Model Dataset Cohort Flowchart

**[Response Ends]**

**sp.17. Describe the denominator exclusions.**

*Brief narrative description of exclusions from the target population.*

**[Response Begins]**

The measure has no exclusions.

**[Response Ends]**

**sp.18. Provide details needed to calculate the denominator exclusions.**

*All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.*

**[Response Begins]**

Not applicable, the measure has no exclusions.

**[Response Ends]**

**sp.19. Provide all information required to stratify the measure results, if necessary.**

*Include the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate. Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format in the Data Dictionary field.*

**[Response Begins]**

Not applicable, the measure is not stratified.

**[Response Ends]**

**sp.20. Is this measure adjusted for socioeconomic status (SES)?**

**[Response Begins]**

No

**[Response Ends]**

**sp.21. Select the risk adjustment type.**

*Select type. Provide specifications for risk stratification and/or risk models in the Scientific Acceptability section.*

**[Response Begins]**

Statistical risk model with risk factors (specify number of risk factors)

**[Statistical risk model with risk factors (specify number of risk factors) Please Explain]**

14

**[Response Ends]**

**sp.22. Select the most relevant type of score.**

*Attachment: If available, please provide a sample report.*

**[Response Begins]**

Rate/proportion

**[Response Ends]**

**sp.23. Select the appropriate interpretation of the measure score.**

*Classifies interpretation of score according to whether better quality or resource use is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score*

**[Response Begins]**

Better quality = Lower score

**[Response Ends]**

**sp.24. Diagram or describe the calculation of the measure score as an ordered sequence of steps.**

*Identify the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period of data, aggregating data; risk adjustment; etc.*

**[Response Begins]**

First, identify the cohort as those patients meeting all inclusion criteria (no exclusion criteria):

- Enrolled in Medicare FFS Part A and Part B at least for 6-months prior (out of the 12-months prior) to TAD/TIP encounter as well as coverage at the time of the intervention and one-month post-intervention.
- Aged 18 years and older.
- Have an encounter with an ambulance provider where the patient was triaged to TAD/TIP. The measure includes all eligible TAD/TIP encounters for each patient, meaning a patient could show up in the cohort multiple times. However, if multiple EMS encounters occur in the same day, choose one encounter.
- Are not enrolled in Medicare hospice at the time of the TAD/TIP encounter with the ambulance provider, with hospice status identified through enrollment in Medicare hospice services in the patient enrollment file.

Second, identify the observed outcome for those patients who had an ED visit or death date within three days of being included in the cohort, removing outcomes from patients who:

- Have an ED visit and are discharged with a primary diagnosis related to mental health and substance use disorder (MH/SUD), listed in [Table 1](#).
- If patients receive multiple TAD/TIP interventions within three days before an ED visit or death, the ED or death *only* counts as an outcome event for the proximal encounter.

Third, calculate the measure score, the risk standardized ED visit rate (RSEDVR), detailed below.

The number of outcome events *predicted* for eligible beneficiaries seen by the ambulance provider given their case mix and the **provider's quality**

The number of outcome events *expected* for eligible beneficiaries seen by the ambulance provider given their case mix and the **average provider quality** in the cohort

**X** The cohort-wide outcome rate

#### Calculation for the Risk Standardized ED Visit Rate

The formula denotes the number of outcome events predicted for eligible patients seen by an ambulance provider given their case mix and the provider's quality DIVIDED BY the number of outcome events *expected* for eligible patients seen by the ambulance provider given their case mix and average provider's quality, MULTIPLIED by the cohort-wide outcome rate. Further details are provided below.

To calculate the measure score, the RSEDVR, a hierarchical generalized linear model (HGLM)-based approach common for CMS quality measures will be used.<sup>1,2</sup> This approach accounts for both clustering of patients within ambulance providers and the variation in patient case-mix across ambulance providers.

In the below equation, let  $Y_{ij}$  denote the presence of the outcome after a TAD/TIP encounter  $i$  by ambulance providers  $j$  ( $Y_{ij}$  is equal to 1 if a patient has an ED visit or dies within three days). We assume the outcome is related linearly to the covariates via a logit function:

$$\text{logit}(\text{Prob}(Y_{ij} = 1)) = \alpha_j + \beta^* Z_{ij}$$

$$\alpha_j = \mu + \omega_j; \omega_j \sim N(0, \tau^2)$$

$$j=1, \dots, J; i=1, \dots, n_j$$

where  $Z_{ij} = (Z_{ij1}, Z_{ij2}, \dots, Z_{ijk})$  is a set of  $k$  encounter-level covariates for the patient at the time of the TAD/TIP encounter;  $J$  denotes the total number of ambulance providers;  $n_j$  denotes the number of index encounters for ambulance providers  $j$ ;  $\alpha_j$  represents the ambulance provider specific intercept;  $\mu$  is the adjusted average intercept over all ambulance providers; and  $\tau^2$  is the between-provider variance components. The HGLM is estimated using the SAS software system (GLIMMIX procedure).

To derive the RSEDVR for provider  $j$ ,  $RSEDVR_j$ , we calculate the predicted number of ED visits and the expected number of ED visits for the ambulance provider. The predicted number of ED visits for each provider is calculated as the sum of the predicted probability of ED visits for each encounter from the HGLM output including the provider specific (random) intercept. The expected number of ED visits for each ambulance provider is similarly calculated as the sum of the predicted probabilities of an ED visit for each encounter with the average intercept. Using the notation of the previous section, the measure score for each ambulance provider,  $RSEDVR_j$ , is calculated as:



$$RSEDVR_j = \text{pred}_j / \text{exp}_j * \bar{y}$$

where

$$\text{pred}_j = \text{logit}^{-1}(\alpha_j + \hat{\beta} * Z_{ij}) \quad (2)$$

$$\text{exp}_j = \text{logit}^{-1}(\mu + \hat{\beta} * Z_{ij}) \quad (3)$$

,  $\hat{\beta}$  represents the estimated coefficients for risk factors, and  $\bar{y}$  is the population outcome rate.

References:

1. AHRQ. Agency For HealthCare and Quality. 2022.
2. National Quality Forum. Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement 2019.

[Response Ends]

**sp.27. If measure testing is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.**

*Examples of samples used for testing:*

- Testing may be conducted on a sample of the accountable entities (e.g., hospital, physician). The analytic unit specified for the particular measure (e.g., physician, hospital, home health agency) determines the sampling strategy for scientific acceptability testing.
- The sample should represent the variety of entities whose performance will be measured. The [2010 Measure Testing Task Force](#) recognized that the samples used for reliability and validity testing often have limited generalizability because measured entities volunteer to participate. Ideally, however, all types of entities whose performance will be measured should be included in reliability and validity testing.
- The sample should include adequate numbers of units of measurement and adequate numbers of patients to answer the specific reliability or validity question with the chosen statistical method.
- When possible, units of measurement and patients within units should be randomly selected.

[Response Begins]

Not applicable. The measure is not based on a sample or survey.

[Response Ends]

**sp.30. Select only the data sources for which the measure is specified.**

[Response Begins]

Claims

Other (specify)

**[Other (specify) Please Explain]**

Medicare Enrollment Data and other administrative data.

**[Response Ends]**

**sp.31. Identify the specific data source or data collection instrument.**

*For example, provide the name of the database, clinical registry, collection instrument, etc., and describe how data are collected.*

**[Response Begins]**

**Data sources for the measure:**

- **Medicare Part A inpatient and Part B outpatient claims:** This data source contains claims data for FFS inpatient and outpatient services including: Medicare inpatient hospital care, outpatient hospital services, as well as inpatient and outpatient physician claims for the 12 months prior to the encounter to determine eligibility for the cohort.
- **Medicare Beneficiary Summary Files (MBSF) and Enrollment Databases:** These datasets contain Medicare beneficiary demographic, benefit/coverage, and vital status information. These data source was used to obtain information on several inclusion/exclusion indicators such as Medicare enrollment in hospice. Mortality outcomes are identified using the Medicare Beneficiary Summary Files (MBSF) in Chronic Condition Warehouse (CCW), using the date of death from CMS Common Medicare Environment (CME).<sup>1</sup>

**Reference:**

1. CODEBOOK: Medicare Beneficiary Summary File (MBSF) Base with Medicare Part A, B, C, and D. Chronic Condition Warehouse. February 2021. Version 1.4.

**[Response Ends]**

**sp.32. Provide the data collection instrument.**

**[Response Begins]**

No data collection instrument provided

**[Response Ends]**

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate fields in the Scientific Acceptability sections of the Measure Submission Form.

- Measures must be tested for all the data sources and levels of analyses that are specified. If there is more than one set of data specifications or more than one level of analysis, contact NQF staff about how to present all the testing information in one form.
- All required sections must be completed.
- For composites with outcome and resource use measures, Questions 2b.23-2b.37 (Risk Adjustment) also must be completed.

- If specified for multiple data sources/sets of specifications (e.g., claims and EHRs), Questions 2b.11-2b.13 also must be completed.
- An appendix for supplemental materials may be submitted (see Question 1 in the Additional section), but there is no guarantee it will be reviewed.
- Contact NQF staff with any questions. Check for resources at the [Submitting Standards webpage](#).
- For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for the [2021 Measure Evaluation Criteria and Guidance](#).

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a. Reliability testing demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For instrument-based measures (including PRO-PMs) and composite performance measures, reliability should be demonstrated for the computed performance score.

2b1. Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure;

AND

If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

2b3. For outcome measures and other measures when indicated (e.g., resource use):

- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; 14,15 and has demonstrated adequate discrimination and calibration
- rationale/data support no risk adjustment/ stratification.

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful 16 differences in performance;

OR

there is evidence of overall less-than-optimal performance.

2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b6. Analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias.

2c. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:

2c1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and

2c2. the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.

(if not conducted or results not adequate, justification must be submitted and accepted)

## Definitions

Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.

Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

Risk factors that influence outcomes should not be specified as exclusions.

With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of \$25 in cost for an episode of care (e.g., \$5,000 v.\$5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:

### Current Submission:

Updated testing information here.

### Previous (Year) Submission:

Testing from the previous submission here.

## 2a.01. Select only the data sources for which the measure is tested.

### [Response Begins]

Claims

Other (specify)

[Other (specify) Please Explain]

[Enrollment Data](#)

### [Response Ends]

**2a.02. If an existing dataset was used, identify the specific dataset.**

*The dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry).*

**[Response Begins]**

The dataset used for testing was the ET3 Model Dataset. Specifically, this dataset included Medicare Parts A and B claims as well as the Medicare Enrollment Database (EDB). To assess social risk factors, we used the Census as well as claims data (dual eligible status obtained through enrollment data; Agency for Healthcare Research and Quality (AHRQ) socioeconomic status (SES) index score calculated from the American Community Survey). The race variable (Black) was obtained through the Medicare Enrollment Database. The datasets used vary by testing type; see Section 2a.07 and 2a.08 for additional details.

**[Response Ends]**

**2a.03. Provide the dates of the data used in testing.**

*Use the following format: "MM-DD-YYYY - MM-DD-YYYY"*

**[Response Begins]**

The dates used vary by testing type; see Section 2a.07 and 2a.08 for details.

**[Response Ends]**

**2a.04. Select the levels of analysis for which the measure is tested.**

*Testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan.*

*Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.*

*Please do not select:*

- *Clinician: Clinician*
- *Population: Population*

**[Response Begins]**

Clinician: Group/Practice

Other (specify)

**[Other (specify) Please Explain]**

Ambulance Providers and Suppliers.

**[Response Ends]**

**2a.05. List the measured entities included in the testing and analysis (by level of analysis and data source).**

*Identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample.*

**[Response Begins]**

Measured entities are ambulance providers (and suppliers), specifically those billing to treat Medicare FFS patients 18 years or older with an ET3 Intervention (TAD/TIP). In analyses included with this NQF submission, there were 46 ambulance providers that provided TAD/TIP interventions.

**[Response Ends]**

**2a.06. Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis), separated by level of analysis and data source; if a sample was used, describe how patients were selected for inclusion in the sample.**

*If there is a minimum case count used for testing, that minimum must be reflected in the specifications.*

**[Response Begins]**

There were 1,410 patients included in the ET3 Model Dataset. For patient characteristics, see [Table 2](#) below. On average, the patients were 74 years of age with a minimum age of 14 and maximum age of 108. Race and ethnicity were grouped into five categories with more than three quarters of the patients being white. In 2019, 1,095 patients did not have dual eligibility versus 315 who did.

Description	N (%)
<b>Age in Measure Year</b>	-
Mean (SD)	74 (14)
Minimum, Maximum	24 (108)
Q2, Interquartile Range (QR)	76 (16)
≥65	1,151 (81.63%)
<65	259 (18.37%)
<b>Sex</b>	-
Male	582 (41.28%)
Female	828 (58.72%)
<b>Race/Ethnicity</b>	-
Unknown	14 (0.99%)
White	1,096 (77.73%)
Black	267 (18.94%)
Other	6 (0.43%)
Asian	6 (0.43%)
Hispanic	17 (1.21%)
North America Native	4 (0.28%)
<b>Dual Eligibility in 2019</b>	-
No	1,095 (77.66%)
Yes	315 (22.34%)

Table 2 ET3 Model Dataset Patient Demographic Information (N= 1,410 patients)

**[Response Ends]**

**2a.07. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing.**

**[Response Begins]**

Measure testing was conducted using the **ET3 Model Dataset**. The ET3 Model Dataset contains claims data submitted by ambulance providers participating in the ET3 Model available in the Chronic Conditions Data Warehouse (CCW) as of August 2022. This data was from ambulance provider carrier claims and short-term acute care hospital facility claims from the CCW, limited to only ambulance providers participating in the ET3 Model. Patients were limited to those meeting inclusion criteria as outlined in section above.

- Dates of data: January 2021 – August 2022
- Number of patients in the dataset: 1,410
- Number of patient encounters in the dataset: 1,552
- Number of measured entities (ambulance providers): 46

A description of the social risk factor datasets including dates of data are shown below in [2a.08](#).

**[Response Ends]**

**2a.08. List the social risk factors that were available and analyzed.**

*For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.*

**[Response Begins]**

We selected socioeconomic status (SES) variables to analyze after reviewing the literature, developing our conceptual model, and examining available national data sources. The causal pathways for SES variable selection are described below in Section 2b.23. The SES variables used for analysis were:

- Dual eligible status: Dual eligible status (i.e., enrolled in both Medicare and Medicaid) patient-level data is obtained from the CMS Master Beneficiary Summary File (MBSF).<sup>1</sup>

Following guidance from ASPE (ASPE 2016; ASPE 2020), NQF (NQF, 2022), and a body of literature demonstrating differential health care and health outcomes among dual eligible patients, we identified dual eligibility as a key variable.<sup>2,3,4</sup> We recognize that Medicare-Medicaid dual eligibility has limitations as a proxy for patients' income or assets because it is a dichotomous variable. We also acknowledge that it is important to test a wider variety of social risk factors, including key variables such as education and poverty level. Therefore, we also provide testing using the AHRQ-SES index score, a validated composite measure of social risk factors based on census data linked to as small a geographic unit as possible, described below.

- AHRQ-validated SES index score includes: percentage of people in the labor force who are unemployed, percentage of people living below poverty level, median household income, median value of owner-occupied dwellings, percentage of people ≥25 years of age with less than a 12th grade education, percentage of people ≥25 years of age completing ≥4 years of college, and percentage of households that average ≥1 people per room.

We selected the AHRQ SES index score because it is a well-validated variable that describes the average SES of people living in defined geographic areas.<sup>5</sup> Its value as a proxy for patient-level information is dependent on having the most granular-level data with respect to communities that patients live in. We considered the area deprivation index (ADI) among many other potential indicators when we initially evaluated the impact of social risk factors. We ultimately did not include the ADI, as coefficients used to derive the ADI had not been updated in recent years at the time of the initial development of the measure. More recently, the coefficients for the ADI have been updated and we resultantly compared the ADI with the AHRQ SES Index and determined that they were highly correlated. In this submission, we present analyses using the census block level, the most granular level possible using

American Community Survey (ACS) data. A census block group is a geographical unit used by the US Census Bureau which is between the census tract and the census block. It is the smallest geographical unit for which the bureau publishes sample data. The target size for block groups is 1,500 and they typically have a population of 600 to 3,000 people. We used 2013-2017 ACS data and mapped patients' 5-digit ZIP codes via vendor software to the census block group level. Given the variation in cost of living across the country, the median income and median property value components of the AHRQ SES Index were adjusted by regional price parity values published by the Bureau of Economic Analysis (BEA). We then calculated an AHRQ SES Index score for census block groups linkable to 9-digit ZIP codes. On a scale of 1-100, we found the lowest quartile of the AHRQ SES Index to be those below or equal to 46, and therefore, 46 is used as the cut-off to divide patients dichotomously as low SES (below or equal to 46) and high SES (above 46).

References:

1. Waldo DR. Accuracy and Bias of Race/Ethnicity Codes in the Medicare Enrollment Database. Health Care Financing Review. 2004;26(2). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4194866/>
2. ASPE 2016: Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation (HHS). Report to Congress: Social Risk factors and Performance Under Medicare's Value-based Payment Programs. 2016; <https://aspe.hhs.gov/pdf-report/report-congress-social-risk-factors-and-performance-under-medicare-value-based-purchasing-programs>. Accessed December 24, 2022.
3. ASPE 2020: Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation (HHS). Report to Congress: Social Risk factors and Performance Under Medicare's Value-based Payment Programs. Second Report to Congress on Social Risk and Medicare's Value-Based Purchasing Programs. 2020. <https://aspe.hhs.gov/reports/second-report-congress-social-risk-medicare-value-based-purchasing-programs>. Accessed December 24, 2022.
4. NQF 2022: National Quality Forum (NQF). Developing and Testing Risk Adjustment Models for Social and Functional Status-Related Risk within Healthcare Performance Measurement.; 2022. <https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=96087>. Accessed December 24, 2022.
5. Bonito A, Bann C, Eicheldinger C, Carpenter L. Creation of new race-ethnicity codes and socioeconomic status (SES) indicators for Medicare beneficiaries. Final Report, Sub-Task. 2008;

**[Response Ends]**

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.09 check patient or encounter-level data; in 2a.010 enter “see validity testing section of data elements”; and enter “N/A” for 2a.11 and 2a.12.

**2a.09. Select the level of reliability testing conducted.**

*Choose one or both levels.*

**[Response Begins]**

Accountable Entity Level (e.g., signal-to-noise analysis)

**[Response Ends]**

**2a.10. For each level of reliability testing checked above, describe the method of reliability testing and what it tests.**

*Describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used.*



**[Response Begins]**

We provide the signal-to-noise reliability statistic (provider-level reliability), which is the reliability with which individual units (ambulance providers) are measured. We analyzed this statistic among all providers and also those with 20 or more encounters, reporting the mean, standard deviation, and median, quartiles, minimum and maximum.

We used the formula presented by Adams and colleagues (2010) to calculate provider-level reliability.<sup>1</sup> In this formula, provider-to-provider variance is estimated from the hierarchical logistic regression model,  $n$  is equal to each provider's observed case size, and the provider error variance is estimated using the variance of the logistic distribution ( $\pi^2/3$ ). The provider-level reliability testing is limited to providers with at least 20 encounters.

Signal-to-noise reliability scores can range from 0 to 1. A reliability of zero implies that all variability in a measure is attributable to measurement error. A reliability of one implies that all variability is attributable to real difference in performance.

Specifically, the signal-to-noise reliability score for agency  $j$ ,  $R_j$  is calculated as:

$$R_j = \frac{n_j ICC}{1 + (n_j - 1) ICC}$$

while

$$ICC = \frac{\tau^2}{\tau^2 + \pi^2/3}$$

$n_j$  is the number of TAD/TIP encounters for Model Participants  $j$ ,  $\tau^2$  is the between agency variance in the HGLM model specified above and represent the signal, and  $\pi^2/3$  represents the noise for a logistic regression.

So,  $R_j$  ranges from 0 to 1.0. The higher the score, the higher reliability. Also, we can see that the reliability of ambulance provider score will vary depending on the number of TAD/TIP encounters. Entities with higher volume will tend to have more reliable scores, while those with lower volume will tend to have fewer reliable scores.

Reference:

1. Adams J, Mehrota, A, Thoman J, McGlynn, E. (2010). Physician cost profiling – reliability and risk of misclassification. NEJM, 362(11): 1014-1021.

**[Response Ends]**

**2a.11. For each level of reliability testing checked above, what were the statistical results from reliability testing?**

*For example, provide the percent agreement and kappa for the critical data elements, or distribution of reliability statistics from a signal-to-noise analysis. For score-level reliability testing, when using a signal-to-noise analysis, more than just one overall statistic should be reported (i.e., to demonstrate variation in reliability across providers). If a particular method yields only one statistic, this should be explained. In addition, reporting of results stratified by sample size is preferred (pg. 18, [NQF Measure Evaluation Criteria](#)).*

**[Response Begins]**

We calculated the signal-to-noise reliability statistic among ambulance providers (Table 3). Among all providers, the median reliability was 0.210 (IQR 0.046-0.615). Among providers with at least 20 encounters, the median reliability was 0.665 (IQR 0.665-0.844).

Statistics			All Providers (N=46)	Providers with 20 + Encounters (N=15)
Number of Encounters			1,552	1,416
Mean (SD)			0.338 (0.297)	0.719 (0.138)
Median (IQR)			0.210 (0.046-0.615)	0.665 (0.615-0.844)

Table 3 ET3 Model Dataset Signal-to-Noise Reliability Results for All Providers and Providers with 20+ Encounters

[Response Ends]

**2a.12. Interpret the results, in terms of how they demonstrate reliability.**

*(In other words, what do the results mean and what are the norms for the test conducted?)*

[Response Begins]

A median reliability statistic of 0.665 for providers with at least 20 encounters (the likely reporting/payment calculation cutoff) indicates sufficient reliability by NQF standards. This statistic provides supporting evidence that there is a true quality difference (signal) between ambulance providers relative to the measurement error (noise). We acknowledge that this result includes a limited number of providers. However, the availability of this new and novel measure will allow for additional data collection in future years.

[Response Ends]

**2b.01. Select the level of validity testing that was conducted.**

[Response Begins]

Systematic assessment of face validity of performance measure score as an indicator of quality or resource use (i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance)

[Response Ends]

**2b.02. For each level of testing checked above, describe the method of validity testing and what it tests.**

*Describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used.*

[Response Begins]

To systematically assess face validity, a Quality Workgroup was convened and composed of 11 members that included measured entities in the voluntary CMMI ET3 Model. Members were selected with diverse experiences, backgrounds, perspectives and involvement in the EMS setting, with selection criteria outlined below:

- Emergency Medical Service (EMS) subject matter experts (SMEs) from diverse backgrounds (e.g., fire/municipal, private-for-profit/non-profit, hospital based, large/small providers, urban/rural, super-rural)
- EMS Medical Directors

- Continuous Quality Improvement (CQI)/Quality Assurance (QA) Managers with direct ET3 Intervention experience
- Non-Participant Quality Oversight SMEs (e.g., individual SMEs from National EMS Quality Alliance [NEMSQA], National Association of EMS Physicians [NAEMSP], National Association of EMS Officials [NASEMSO]).

Quality Workgroup Members were provided an information sheet ahead of time about the measure, informed about the measure specifications and rationale during a live session, encouraged to make suggestions, and asked the following questions:

- Do you believe the measure, as specified, can be used to distinguish between better or worse quality of care among ambulance providers?
  - Responses were limited to one of the following: strongly agree, somewhat agree, somewhat disagree, strongly disagree.
- How do you think this measure will provide useful information for providers and please provide rationale?

All Quality Workgroup members responded to the above questions. Of relevance to face validity, Quality Workgroup members rated the ability of the measure to help distinguish better and worse quality of care of ambulance providers.

#### [Response Ends]

#### 2b.03. Provide the statistical results from validity testing.

*Examples may include correlations or t-test results.*

#### [Response Begins]

9 out of 11 of quality workgroup members (82%) strongly agreed or somewhat agreed that the Post Triage ED Visit Rate Measure can be used to distinguish better or worse quality of care among ambulance providers.

Statements - Respondents	Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree
<b>Statement 1: Importance – TEP</b>	3	6	2	0

Table 4 Face Validity Results Distinguishing Quality of Care of Ambulance Providers

Among Quality Workgroup members who agreed the measure exhibits face validity, one stakeholder stated that the risk adjustment and overall measure calculation was well thought out. Another stakeholder agreed that the measure can determine where quality improvement can be assessed and be used to improve the quality or standard of care provided by ambulance providers. Additionally, stakeholders agreed that the Risk Adjusted Post-Ambulance Provider Triage ED Visit Rate Measure will provide useful information to ambulance providers and to CMS. Several Quality Workgroup members stated that this measure would provide beneficial information to ambulance providers to identify provider education effectiveness and triage appropriateness, acknowledging the correlation between the assessment capability of a given provider and the subsequent outcome of a given patient. With patient safety being of paramount concern, this measure allows ambulance providers to determine whether the TAD/TIP encounter they provided was clinically appropriate and did not result in an ED visit or death within 3 days.

No members of the Workgroup selected “strongly disagree.”

Among the 2 people who somewhat disagreed, one person praised the risk adjustment and measure calculation but suggested a potential need to account for ED visits ‘related’ to the initial triage chief complaint or potentially providing the ED discharge diagnosis to ambulance providers, so they are able to evaluate further themselves. We agree with the commenter that aggregate data regarding the ED discharge diagnosis and associated ambulance provider diagnosis should be provided, and that request will be considered once the measure is implemented.

Another person who selected ‘somewhat disagree’ noted that the measure captured the quality of triage, but perhaps not the level of care provided during the TAD/TIP intervention.

[Response Ends]

**2b.04. Provide your interpretation of the results in terms of demonstrating validity. (i.e., what do the results mean and what are the norms for the test conducted?)**

[Response Begins]

Quality Workgroup stakeholders, which included measured entities, strongly supported the face validity of the measure and its inclusion in CMMI’s voluntary payment model.

- Stakeholders (inclusive of measured entities) raised no major threats to measure validity.
- Stakeholders (inclusive of measured entities) raised no concerns about the adequacy of risk adjustment.
- Stakeholders (inclusive of measured entities) raised no concerns around the construct of the measure score.

We agree with feedback from commenters, and reiterate this measure is a first step towards identifying patient safety in triage decisions.

[Response Ends]

**2b.05. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified.**

*Describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided in Importance to Measure and Report: Gap in Care/Disparities.*

[Response Begins]

Examination of ambulance provider-level results include measure scores for all ambulance providers shown as a risk standardized ED visit rate (RSEDVR) and those with at least 20 encounters. We present summary statistics including the mean (SD), median (IQR), and the minimum (min) and maximum (max).

[Response Ends]

**2b.06. Describe the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities.**

*Examples may include number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined.*

[Response Begins]

Our results show a wide variation in measure scores. As shown in Table 5, 46 ambulance providers contributed 1,552 TAD/TIP encounters, with a median measure score and IQR of 19.91% (19.15-22.15%). Ambulance providers with at least 20 TAD/TIP encounters had similar median measure scores and IQR of 21.57% (17.67-23.03%). Among ambulance providers with at least 20 TAD/TIP encounters, we observed wide variation in measure score performance with the range being from 12.33%- 25.72%.

Statistics	All Providers (N=46)	Providers with 20+ Encounters (N=15)
Number of Encounters	1,552	1,416

Statistics	All Providers (N=46)	Providers with 20+ Encounters (N=15)
Mean (SD)	20.62% (3.25%)	20.20% (3.62%)
Median (IQR)	19.91% (19.15- 22.15%)	21.57% (17.67- 23.03%)
Range (min. – max.)	12.33%- 33.05%	12.33%- 25.72%

Table 5 ET3 Summary Statistics of Measure Score, Risk Standardized ED Visit Rate (RSEDVR), for All Providers and Providers with 20 or More Patients, ET3 Model Dataset January 2021 – April 2022

[Response Ends]

**2b.07. Provide your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities.**

*In other words, what do the results mean in terms of statistical and meaningful differences?*

[Response Begins]

Among ambulance providers with at least 20 TAD/TIP encounters, we observed wide (2-fold) variation in measure scores, indicating a possible opportunity for quality improvement. These early estimates of performance variation are conservative as they reflect measured entities that are early adopters of the ET3 model and likely very selective in the delivery TIP or TAD Interventions. Therefore, the quality of care may vary substantially as the measure is implemented and more ambulance providers meets the minimum case count requirements and adapt their clinical processes.

[Response Ends]

**2b.08. Describe the method of testing conducted to identify the extent and distribution of missing data (or non-response) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders). Include how the specified handling of missing data minimizes bias.**

*Describe the steps—do not just name a method; what statistical analysis was used.*

[Response Begins]

The three-day Post Triage ED Visit Rate Measure used claims-based data for development and testing. There was no missing data in the claims-based development and testing data.

[Response Ends]

**2b.09. Provide the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data.**

*For example, provide results of sensitivity analysis of the effect of various rules for missing data/non-response. If no empirical sensitivity analysis was conducted, identify the approaches for handling missing data that were considered and benefits and drawbacks of each).*

[Response Begins]

The three-day Post Triage ED Visit Rate Measure used claims-based data for development and testing. There was no missing data in the claims-based development and testing data.

[Response Ends]

**2b.10. Provide your interpretation of the results, in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders), and how the specified handling of missing data minimizes bias.**

*In other words, what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis was conducted, justify the selected approach for missing data.*

**[Response Begins]**

Not applicable. There was no missing data in the claims-based development and testing data.

**[Response Ends]**

Note: This item is directed to measures that are risk-adjusted (with or without social risk factors) OR to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eQMs). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator). Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.

**2b.11. Indicate whether there is more than one set of specifications for this measure.**

**[Response Begins]**

No, there is only one set of specifications for this measure

**[Response Ends]**

**2b.12. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications.**

*Describe the steps—do not just name a method. Indicate what statistical analysis was used.*

**[Response Begins]**

**[Response Ends]**

**2b.13. Provide the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications.**

*Examples may include correlation, and/or rank order.*

**[Response Begins]**

**[Response Ends]**

**2b.14. Provide your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications.**

*In other words, what do the results mean and what are the norms for the test conducted.*

**[Response Begins]**

**[Response Ends]**

**2b.15. Indicate whether the measure uses exclusions.**

**[Response Begins]**

N/A or no exclusions

**[Response Ends]**

**2b.16. Describe the method of testing exclusions and what was tested.**

*Describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used?*

**[Response Begins]**

There are no denominator exclusions for this measure. By design, populations for whom ambulance service provider outcomes may be more challenging to link to triage quality (e.g., beneficiaries receiving hospice services, beneficiaries with mental health conditions or substance use disorders, etc.) were not excluded from the quality measure (denominator); instead, outcome events for these populations were removed from the numerator (See sp.14).

**[Response Ends]**

**2b.17. Provide the statistical results from testing exclusions.**

*Include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores.*

**[Response Begins]**

Not applicable. There were no exclusions for this measure.

**[Response Ends]**

**2b.18. Provide your interpretation of the results, in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results.**

*In other words, the value outweighs the burden of increased data collection and analysis. Note: If patient preference is an exclusion, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion.*

**[Response Begins]**

Not applicable. There were no exclusions for this measure.

**[Response Ends]**

**2b.19. Check all methods used to address risk factors.**

**[Response Begins]**

Statistical risk model with risk factors (specify number of risk factors)

**[Statistical risk model with risk factors (specify number of risk factors) Please Explain]**

There were 14 risk factors included in the model.

Other approach to address risk factors (specify)

**[ Other approach to address risk factors (specify) Please Explain]**

The expanded risk model of 30 risk variables was tested in the Development Sample of the ET3 Model Dataset. The risk model was then reduced to fewer variables for use with lower volumes of data by clinically grouping certain risk variables together. Additionally, two variables representing mental health and substance-use disorder were removed to account for the fact that ED visits with a primary diagnosis related to MH/SUD are not counted as a measure outcome. The candidate social risk variables were then tested in the ET3 Model Dataset, detailed in [section 2b.23](#).

**[Response Ends]**

**2b.20. If using statistical risk models, provide detailed risk model specifications, including the risk model method, risk factors, risk factor data sources, coefficients, equations, codes with descriptors, and definitions.**

**[Response Begins]**

The goal of risk adjustment is to account for differences between ambulance providers in patient demographic and clinical characteristics that are potentially related to the outcome but are unrelated to quality of care. This measure risk adjusts to account for factors that are associated with the outcome (ED visit or death within three days) that vary across ambulance providers and are unrelated to quality of care. Accounting for case-mix differences is important as certain ambulance providers care for an older and more comorbid patient population potentially more likely to have a post-TAD/TIP encounter ED visit even if properly triaged. Through the risk-adjustment modeling, a higher expected outcome rate is set for ambulance providers caring for patients with a higher case mix.

We first identified candidate variables conceptually, through a literature review, environmental scan, clinical and expert input. We then validated the initially selected variables based on an analysis of ED utilization using a large dataset of Medicare beneficiaries. Through use of this dataset we identified an expanded list of 30 risk adjustment variables, including clinical and demographic (age) factors. We then applied the risk model to the ET3 Model Dataset, and further reduced it to include 14 risk adjustment variables to accommodate for the limited number of TAD/TIP encounters. The risk model was finalized as outlined below and tested using the ET3 Model Dataset. All risk model and measure score results are from the ET3 Model Dataset. The final variables are presented in the table below.

**Candidate Clinical and Demographic Risk Variables:**

We considered clinical medical history (comorbidities, frailty, etc.) and age as candidate variables:

- Patient comorbidities for inclusion in risk adjustment were identified through inpatient and outpatient administrative claims during the six months prior to entering the cohort.
- We align with other CMS outcome measures by using the Yale-Modified FY20 v24 CC Map which is derived from the publicly available CMS condition categories (CMS-CCs) to group ICD-10 diagnosis codes into CMS-CCs. We selected comorbidities based on clinical relevance and statistical significance.

The process of testing candidate risk variables for the model included:

- Using the Yale-Modified FY20 v24 CC Map, we examined all CMS-CCs to assess the frequency of each comorbidity and bivariate associations with the outcome with odds ratios.
- We then grouped clinically and statistically similar CMS-CCs together. To alleviate the burden of yearly reevaluation, we align efforts with the NQF-endorsed (#2888) Merit-Based Incentive Payment System



multiple chronic conditions (MIPS MCC) measure. More information about MIPS MCC measures can be found [on CMS website](#).

- similarly covers a broad population, groups conditions together in a clinically and statistically sensible manner, and is [NQF-endorsed](#).
  - We removed CMS-CCs with no clinical relevance to the outcome.
  - Expert clinician review was completed to remove clinically irrelevant candidate variables.
  - Age was added as a categorical variable.
  - This resulted in 60 clinical candidate risk variables in addition to age.
- Encounters were split into two groups: the Development Sample and the Validation Sample. We then performed stepwise model selection using logistic regression to identify and retain statistically significant risk variables.
- This step resulted in the retention of 30 risk factors (groups of CMS-CCs as comorbidities), including age.
  - The expanded risk model of 30 risk variables was tested in the Development Sample of the ET3 Model Dataset. The risk model was then reduced to fewer variables for use with lower volumes of data by clinically grouping certain risk variables together. Additionally, two variables representing mental health and substance-use disorder were removed to account for the fact that ED visits with a primary diagnosis related to MH/SUD are not counted as a measure outcome. The candidate social risk variables were then tested in the ET3 Model Dataset, detailed in [section 2b.23](#).

Risk Factor
<b>Age</b> (categorical as: 18-65; 66-75; and 76+)
<b>Chronic Obstructive Pulmonary Disease and Asthma</b> (CC 111, 112, 113, 118), <b>Pleural effusion/pneumothorax</b> (CC117), <b>Pneumonia</b> (CC114, 115, 116)
<b>Congestive Heart Failure</b> (CC85), <b>Vascular or circulatory disease</b> (CC106, 107, 108,109)
<b>Dialysis Status</b> (CC 134), <b>Disorders of Fluid/Electrolyte/Acid-Base Balance</b> (CC 24), <b>Urinary Obstruction and Retention</b> (CC142)
<b>Gastrointestinal disease</b> (CC31, 32, 33, 35, 36), <b>Pancreatic disease</b> (CC 34)
<b>Head Injury</b> (CC 166, 167, 168)
<b>Hematological diseases</b> (CC 46, 48), <b>Iron deficiency anemia</b> (CC 49)
<b>Hypertension</b> (CC95), <b>Hypertensive Heart Disease</b> (CC 94), <b>Ischemic heart disease</b> (CC86, 87, 88, 89)
<b>Marked disability/frailty</b> (CC21, 70, 71, 73, 157, 158, 159, 160, 161, 189, 190)
<b>Pelvic Inflammatory Disease and Other Specified Female Genital Disorders</b> (CC 147), <b>Pregnancy</b> (CC150, 151, 152,153, 155, 156)
<b>Septicemia/shock</b> (CC2)
<b>Advanced cancer</b> (CC 8, 9, 10, 13)

Risk Factor
Advanced liver disease (CC27, 28, 29, 30)
Cellulitis, Local Skin Infection (CC 164), Bone/joint/muscle infections/ necrosis (CC 39,40, 41, 42)

Table 6 Final Risk Factors Used for Testing and Measure Implementation

**[Response Ends]**

**2b.21. If an outcome or resource use measure is not risk-adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (i.e., case mix) is not needed to achieve fair comparisons across measured entities.**

**[Response Begins]**

Not applicable.

**[Response Ends]**

**2b.22. Select all applicable resources and methods used to develop the conceptual model of how social risk impacts this outcome.**

**[Response Begins]**

Published literature

**[Response Ends]**

**2b.23. Describe the conceptual and statistical methods and criteria used to test and select patient-level risk factors (e.g., clinical factors, social risk factors) used in the statistical risk model or for stratification by risk.**

*Please be sure to address the following: potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of  $p < 0.10$  or other statistical tests; correlation of  $x$  or higher. Patient factors should be present at the start of care, if applicable. Also discuss any “ordering” of risk factor inclusion; note whether social risk factors are added after all clinical factors. Discuss any considerations regarding data sources (e.g., availability, specificity).*

**[Response Begins]**

Methods for identifying the clinical risk variables in the model are detailed above in Section 2b.20.

In testing for social risk variables, we included in the measure’s conceptual model ways in which social risk factors may influence ED visit rates and the ability of ambulance providers to mitigate these risks. The graphic below describes a conceptual model for ED visitation and includes factors not related to the quality of care provided by the ambulance provider that can potentially influence the decision to offer TAD/TIP or the patient’s outcome at three days. We developed the conceptual model using both published literature, and expert input, described below.

There have been few studies that have examined patient-level social factors associated with ED visits following triage to a non-hospital setting, therefore our conceptual model includes evidence from literature examining social risk factors and a hospital admission following ED treat and discharge.<sup>1</sup>

- Rural residency: The only study to examine the impact of social risk factors on the outcome of an ED visit following triage to a non-ED setting (in Finland), found that living in a rural area was not significantly associated with a subsequent ED visit; patients in rural settings were more likely, however, to experience a primary care visit.<sup>2</sup>

- Income: Patients in the United States with lower income were more likely to re-visit the ED following treat-and-discharge from the ED following an injury, suggesting that low income could also play a role in an ED visit following triage to a non-ED setting.<sup>3</sup>
- Patients of a race other than white were less likely to re-visit the ED following treat-and-discharge from the ED following an injury. Similarly, in a study of outcomes following ED discharge in Medicare patients, black patients were less likely to have a follow-up ambulatory care visit after ED discharge, which in turn was associated with a lower ED visit rate.<sup>4</sup>

Conceptually, many factors could influence the outcome of an ED visit or death within 3 days. Patient comorbidities and age influence the outcome, while there are also mediators such as a patient's access to, or knowledge of, unscheduled care services, which can increase the risk of the outcome if these considerations are lacking. Finally, a patient's demographics, such as their social risk and living situation, may conceptually influence the outcome. If a patient has a live-in caretaker or visiting nurse, this knowledge may influence the ambulance provider's triage decision.

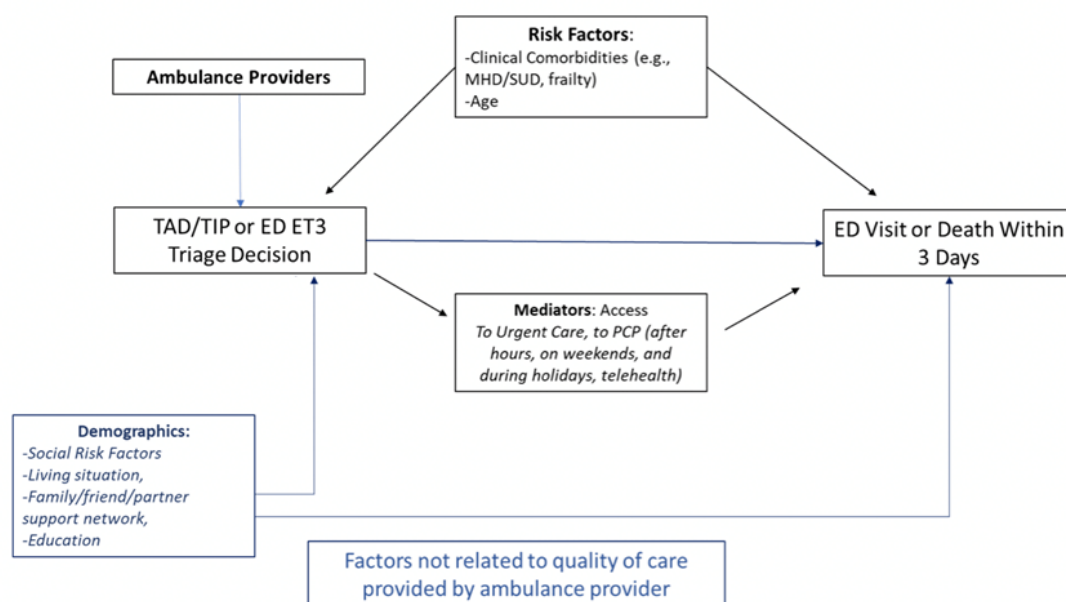


Figure 3 ET3 Conceptual Model of Impact of Social Risks

Using the above conceptual model, we identified two social risk factors for analysis based on a lower acuity, population-wide cohort, and availability of data: a) Dual eligibility in Medicare & Medicaid; and b) AHRQ SES Index. Dual eligibility is often used as an indicator for patients with a high prevalence of chronic conditions and disabilities, substantial care needs, and high health care utilization and costs. This variable is a claims-based and reliable variable. The Agency for Healthcare Research and Quality (AHRQ) has created an index assessing the socioeconomic status of geographic areas using a variety of factors using data from the American Community Survey.<sup>5</sup>

Other factors in the 'Demographics' box may have an influence on the. However, they are more difficult to capture in any database and difficult to measure. Living situation, support network, and education are more likely to influence the original triage decision by the ambulance provider. The influence on the outcome of these variables was not tested due to lack of reliable data available. These factors will be considered during measure reevaluation, pending the availability of reliable sources of data.

Variable	Description	Data level
Dual Eligibility Status: Yes	Dual-eligible for Medicare and Medicaid vs. Medicare-only (reference variable)	Beneficiary
AHRQ SES Index 1: Lowest economic status	Lowest AHRQ quartile for socioeconomic status indicator (higher score = less social risk) vs. other quartiles (reference variable)	Zip code

Table 7 Candidate Social Risk Factors

References:

1. Paulin J, Kurola J, Koivisto M, Irola T. EMS non-conveyance: A safe practice to decrease ED crowding or a threat to patient safety? BMC Emerg Med. 2021 Oct 9;21(1):115. Doi: 10.1186/s12873-021-00508-1. PMID: 34627138; PMCID: PMC8502399.
2. Paulin J, Kurola J, Koivisto M, Irola T. EMS non-conveyance: A safe practice to decrease ED crowding or a threat to patient safety? BMC Emerg Med. 2021 Oct 9;21(1):115. Doi: 10.1186/s12873-021-00508-1. PMID: 34627138; PMCID: PMC8502399.
3. Earl-Royal EC, Kaufman EJ, Hanlon AL, Holena DN, Rising KL, Kit Delgado M. Factors associated with hospital admission after an emergency department treat and release visit for older adults with injuries. Am J Emerg Med. 2017 Sep;35(9):1252-1257. Doi: 10.1016/j.ajem.2017.03.051. Epub 2017 Mar 21. Erratum in: Am J Emerg Med. 2018 Mar 13;: PMID: 28410919; PMCID: PMC5854494.
4. Hing E, Rui P. Emergency department use in the country's five most populous states and the total United States, 2012. National Center for Health Statistics. Published June 2016. Accessed September 10, 2020. <https://www.cdc.gov/nchs/products/databriefs/db252.htm>
5. AHRQ. Agency For HealthCare and Quality. 2022.

[Response Ends]

**2b.24. Detail the statistical results of the analyses used to test and select risk factors for inclusion in or exclusion from the risk model/stratification.**

[Response Begins]

Results from analytic testing are shown below, with parameter estimates and odds ratios shown for demographic and clinical risk variables and their association with the outcome.

Risk Variable	Risk Factor Prevalence (%)	Parameter Estimates (Standard Error)	Odds Ratio (LOR-UOR)	P value
Age 18-65	22%	-0.366 (0.1866)	0.693 (0.481-1)	0.0498
Age 66-75	30%	0.009 (0.153)	1.009 (0.747-1.362)	0.9539
Age 76+	49%	Ref ()	(-)	
Congestive Heart Failure; Vascular or circulatory disease	56%	0.104 (0.163)	1.110 (0.806-1.528)	0.5227
Dialysis Status; Disorders of Fluid/Electrolyte/Acid-Base Balance; Urinary Obstruction and Retention	41%	0.240 (0.159)	1.271 (0.930-1.737)	0.1329
Gastrointestinal disease; Pancreatic disease;	23%	0.045 (0.166)	1.046 (0.756-1.448)	0.7836

Risk Variable	Risk Factor Prevalence (%)	Parameter Estimates (Standard Error)	Odds Ratio (LOR-UOR)	P value
Head Injury	12%	0.372 (0.196)	1.450 (0.988-2.129)	0.0575
Hematological diseases; Iron deficiency anemia	44%	-0.087 (0.160)	0.917 (0.669-1.255)	0.5874
Hypertension; Hypertensive Heart Disease; Ischemic heart disease	76%	-0.052 (0.185)	0.949 (0.660-1.364)	0.7767
Marked disability/frailty	19%	0.177 (0.1723)	1.194 (0.852-1.674)	0.3034
Pelvic Inflammatory Disease and Other Specified Female Genital Disorders; Pregnancy	3%	0.462 (0.369)	1.587 (0.77-3.27)	0.2104
Septicemia/shock	10%	0.399 (0.214)	1.490 (0.979-2.268)	0.0630
Advanced cancer	6%	-0.028 (0.267)	0.973 (0.576-1.642)	0.9177
Advanced liver disease	4%	-0.131 (0.336)	0.878 (0.454-1.697)	0.6978
Bone/joint/muscle infections/necrosis; Cellulitis, Local Skin Infection	49%	-0.126 (0.142)	0.881 (0.667-1.165)	0.3751
Chronic Obstructive Pulmonary Disease and Asthma; Pleural effusion/pneumothorax; Pneumonia	51%	0.148 (0.149)	1.160 (0.865-1.554)	0.3209

Table 8 Risk Model Variable Frequencies, Parameter Estimates and Odds Ratio Estimates Using Logistic Regression Model, ET3 Model Dataset (N=1,552 encounters)

**[Response Ends]**

**2b.25. Describe the analyses and interpretation resulting in the decision to select or not select social risk factors.**

*Examples may include prevalence of the factor across measured entities, availability of the data source, empirical association with the outcome, contribution of unique variation in the outcome, or assessment of between-unit effects and within-unit effects. Also describe the impact of adjusting for risk (or making no adjustment) on providers at high or low extremes of risk.*

**[Response Begins]**

Because of the limited amount of data available in this early, voluntary reporting dataset, there are too few patients with social risk factors in the ET3 dataset to completely evaluate the relationship between social risk factors and other variables in the risk model. While our conceptual model ([Figure 3](#)) considers social risk factors (e.g., dual eligibility, low AHRQ SES) not as part of the initial triage decision but as mediators around access to non-ED care, those variables may interact with existing variables in the risk model in unpredictable ways, and it is premature, given the lack of data for testing, to include them in the statistical model. As participation in the ET3

model increases, we will continue to examine the relationship between social risk factors and the outcome, including the impact on ambulance provider measure scores, to ensure that providers with a higher proportion of patients with social risk factors are not unfairly characterized.

**Analyses:**

To understand the relationship between social risk factors, the outcome (ED visits and death), and the impact on measure scores, we first examined the prevalence of each social risk factor ( low AHRQ SES, and dual eligibility) among patients in the ET3 Model dataset.

Description	Number of Patients (Percent)
Total number of patients in ET3 Model dataset	1,410 (100%)
<b>Dual Eligibility in 2019</b>	-
No	1,095 (77.66%)
Yes	315 (22.34%)
<b>AHRQ SES variable</b>	-
Non-low-AHRQ SES (>46)	1,062 (75.48%)
Low AHRQ SES (<=46)	345 (24.52%)

Table 9 Prevalence of social risk factor among patients in the ET3 Model dataset

We found that there were only 315 patients with dual eligibility status and 345 patients with low AHRQ SES status. While the proportion of patients with social risk factors is substantial, the total number of patients with social risk factors in this voluntary ET3 Dataset is too few to provide reliable or valid assessments of the interaction of these variables with other variables in the risk model, or an assessment of the impact of adjustment on measure scores. The developer will continue to evaluate the impact of social risk factors as additional data becomes available.

**[Response Ends]**

**2b.26. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (describe the steps—do not just name a method; what statistical analysis was used). Provide the statistical results from testing the approach to control for differences in patient characteristics (i.e., case mix) below. If stratified ONLY, enter “N/A” for questions about the statistical risk model discrimination and calibration statistics.**

*Validation testing should be conducted in a data set that is separate from the one used to develop the model.*

**[Response Begins]**

We computed two summary statistics for assessing model performance.<sup>1</sup>

(1) Area under the receiver operating characteristic (ROC) curve (the c-statistic) is the probability that predicting the outcome is better than chance, which is a measure of how accurately a statistical model is able to distinguish between a patient with and without an outcome)

(2) Predictive ability (discrimination in predictive ability measures the ability to distinguish high-risk subjects from low-risk subjects; therefore, we would hope to see a wide range between the lowest decile and highest decile.)

Reference:

1. Harrell FE and Shih YC. Using full probability models to compute probabilities of actual interest to decision makers, Int. J. Technol. Assess. Health Care 17 (2001), pp. 17–26.

**[Response Ends]**

**2b.27. Provide risk model discrimination statistics.**

*For example, provide c-statistics or R-squared values.*

**[Response Begins]**

The **C-statistic** indicated acceptable model discrimination, with a value of 0.601.

We also examined model performance inclusive of an ambulance provider random effect using hierarchical generalized linear models (HGLM). In this model, we identified a c-statistic of 0.690, supporting findings of large variation in performance between ambulance providers and demonstrating that the model can capture patient-level risk as well as variation between ambulance providers. This large increase (about 0.1) in the c-statistic when including the ambulance provider random intercept (effect) explains the lower patient-level c-statistic of 0.601 and indicates that variation is due to provider performance.

**Predictive ability** (lowest decile %, highest decile %) = (12.1, 31.6)

**[Response Ends]**

**2b.28. Provide the statistical risk model calibration statistics (e.g., Hosmer-Lemeshow statistic).**

**[Response Begins]**

We cannot currently provide overfitting calibration statistics given the small sample size of the early testing dataset. However, below in section 2b.29, we present an encounter-level calibration plot using the predicted probability deciles which exhibited acceptable calibration.

**[Response Ends]**

**2b.29. Provide the risk decile plots or calibration curves used in calibrating the statistical risk model.**

*The preferred file format is .png, but most image formats are acceptable.*

**[Response Begins]**

[Figure 4](#) shows the encounter-level calibration plot using the predicted probability deciles. The X-axis is the average predicted probability for an outcome in each decile, and the Y-axis is the observed outcome rate. The vertical lines represent the confidence intervals of the estimate. The results showing deciles of both underprediction and overprediction are likely influenced by the low volume of data.

At one, we unexpectedly find, due to the small sample size, anticipate improvement with more denominator cases.

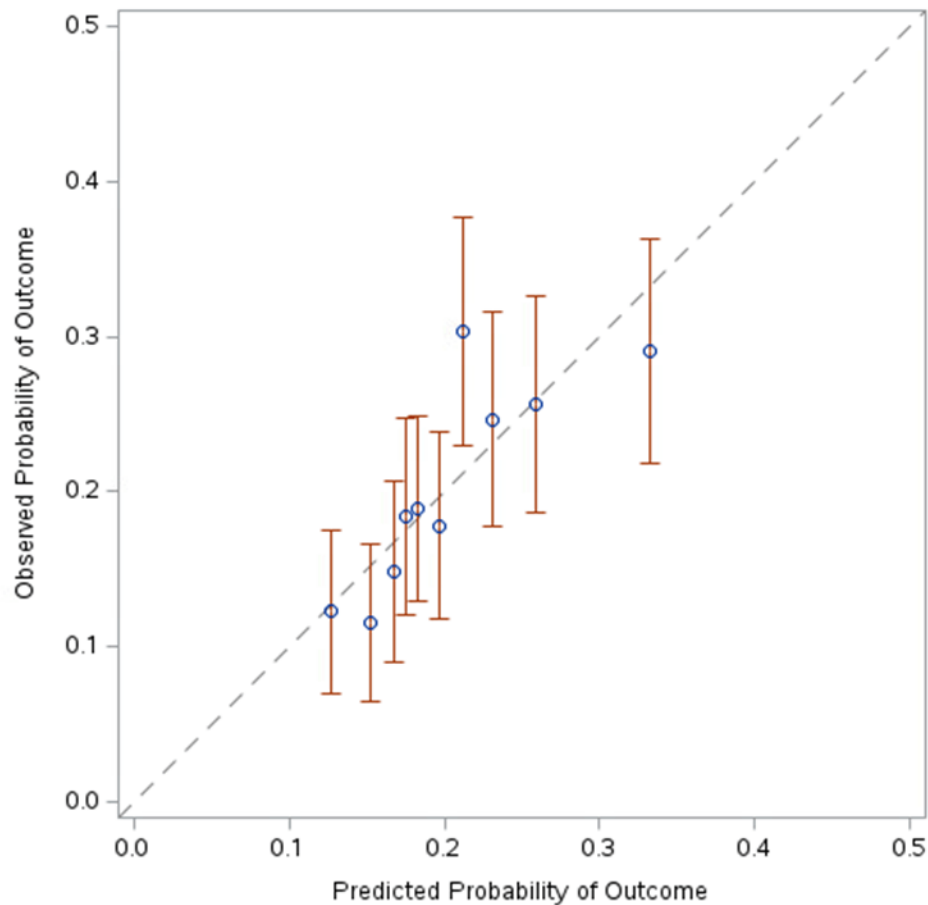


Figure 4 Observed Probability of Outcome vs. Predicted Probability of Outcome – Patient-Level Logistic Regression (ET3 Model Dataset)

[Response Ends]

**2b.30. Provide the results of the risk stratification analysis.**

[Response Begins]

This measure is not stratified.

[Response Ends]

**2b.31. Provide your interpretation of the results, in terms of demonstrating adequacy of controlling for differences in patient characteristics (i.e., case mix).**

*In other words, what do the results mean and what are the norms for the test conducted?*

[Response Begins]

The c-statistic indicated acceptable model discrimination, with a value of 0.601.

As noted earlier, we also examined model performance inclusive of an ambulance provider random effect using hierarchical generalized linear models (HGLM). In this model, we identified a c-statistic of 0.690, supporting findings of large variation in performance between ambulance providers and demonstrating that the model can



capture patient-level risk as well as variation between ambulance providers. This large increase (about 0.1) in the c-statistic when including the ambulance provider random intercept (effect) explains the lower patient-level c-statistic of 0.601 and indicates that variation is due to provider performance.

The model has acceptable calibration, as shown by the decile plots. The results showing deciles of both underprediction and overprediction are likely influenced by the low volume of data.

There was a wide range of observed outcome rates between the lowest and highest predicted probability decile (12.1% vs. 31.6%), which indicates good model discrimination and calibration. The model indicated a wide range between the lowest decile and highest decile, indicating the ability to distinguish high-risk subjects from low-risk subjects.

**[Response Ends]**

**2b.32. Describe any additional testing conducted to justify the risk adjustment approach used in specifying the measure.**

*Not required but would provide additional support of adequacy of the risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed.*

**[Response Begins]**

No additional information.

**[Response Ends]**

### 3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

---

**3.01. Check all methods below that are used to generate the data elements needed to compute the measure score.**

[Response Begins]

[Response Ends]

**3.02. Detail to what extent the specified data elements are available electronically in defined fields.**

*In other words, indicate whether data elements that are needed to compute the performance measure score are in defined, computer-readable fields.*

[Response Begins]

[Response Ends]

**3.03. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using data elements not from electronic sources.**

[Response Begins]

[Response Ends]

**3.04. Describe any efforts to develop an eCQM.**

[Response Begins]

[Response Ends]

**3.06. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

[Response Begins]

[Response Ends]

Consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

**3.07. Detail any fees, licensing, or other requirements to use any aspect of the measure as specified (e.g., value/code set, risk model, programming code, algorithm),**

**Attach the fee schedule here, if applicable.**

[Response Begins]

[Response Ends]



## 4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

---

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making.

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement, in addition to demonstrating performance improvement.

**4a.01. Check all current uses. For each current use checked, please provide:**

- **Name of program and sponsor**
- **URL**
- **Purpose**
- **Geographic area and number and percentage of accountable entities and patients included**
- **Level of measurement and setting**

[Response Begins]

[Response Ends]

**4a.02. Check all planned uses.**

[Response Begins]

[Response Ends]

**4a.03. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing), explain why the measure is not in use.**

*For example, do policies or actions of the developer/steward or accountable entities restrict access to performance results or block implementation?*

[Response Begins]

[Response Ends]

**4a.04. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes: used in any accountability application within 3 years, and publicly reported within 6 years of initial endorsement.**

*A credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*

[Response Begins]

[Response Ends]

**4a.05. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.**

*Detail how many and which types of measured entities and/or others were included. If only a sample of measured entities were included, describe the full population and how the sample was selected.*

[Response Begins]

[Response Ends]

**4a.06. Describe the process for providing measure results, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.**

[Response Begins]

[Response Ends]

**4a.07. Summarize the feedback on measure performance and implementation from the measured entities and others. Describe how feedback was obtained.**

[Response Begins]

[Response Ends]

**4a.08. Summarize the feedback obtained from those being measured.**

[Response Begins]

[Response Ends]

**4a.09. Summarize the feedback obtained from other users.**

[Response Begins]

[Response Ends]

**4a.10. Describe how the feedback described has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.**

[Response Begins]

[Response Ends]

**4b.01. You may refer to data provided in Importance to Measure and Report: Gap in Care/Disparities, but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included). If no improvement was demonstrated, provide an explanation. If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**

[Response Begins]

[Response Ends]

**4b.02. Explain any unexpected findings (positive or negative) during implementation of this measure, including unintended impacts on patients.**

[Response Begins]

[Response Ends]

**4b.03. Explain any unexpected benefits realized from implementation of this measure.**

[Response Begins]

[Response Ends]

## 5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

---

If you are updating a maintenance measure submission for the first time in MIMS, please note that the previous related and competing data appearing in question 5.03 may need to be entered in to 5.01 and 5.02, if the measures are NQF endorsed. Please review and update questions 5.01, 5.02, and 5.03 accordingly.

**5.01. Search and select all NQF-endorsed related measures (conceptually, either same measure focus or target population).**

**NOTE: If there are no related measures, please select N/A.**

*(Can search and select measures.)*

[Response Begins]

[Response Ends]

**5.02. Search and select all NQF-endorsed competing measures (conceptually, the measures have both the same measure focus and target population).**

**NOTE: If there are no competing measures, please select N/A.**

*(Can search and select measures.)*

[Response Begins]

[Response Ends]

**5.03. If there are related or competing measures to this measure, but they are not NQF-endorsed, please indicate the measure title and steward.**

[Response Begins]

[Response Ends]

**5.04. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s), indicate whether the measure specifications are harmonized to the extent possible.**

[Response Begins]

[Response Ends]

**5.05. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**

[Response Begins]

[Response Ends]

**5.06. Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality). Alternatively, justify endorsing an additional measure.**

*Provide analyses when possible.*

[Response Begins]

[Response Ends]

## Appendix

Supplemental materials may be provided in an appendix.:

## Contact Information

**Measure Steward (Intellectual Property Owner):** Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (CORE)

**Measure Steward Point of Contact:** Gettel, Cameron, [cameron.gettel@yale.edu](mailto:cameron.gettel@yale.edu)

Peter, Doris, [doris.peter@yale.edu](mailto:doris.peter@yale.edu)

Attanasio, Sarah, [sarah.attanasio@yale.edu](mailto:sarah.attanasio@yale.edu)

Venkatesh, Arjun, [arjun.venkatesh@yale.edu](mailto:arjun.venkatesh@yale.edu)

**Measure Developer if different from Measure Steward:** Centers for Medicare & Medicaid Services

**Measure Developer Point(s) of Contact:** Gettel, Cameron, [cameron.gettel@yale.edu](mailto:cameron.gettel@yale.edu)

Peter, Doris, [doris.peter@yale.edu](mailto:doris.peter@yale.edu)

Attanasio, Sarah, [sarah.attanasio@yale.edu](mailto:sarah.attanasio@yale.edu)

Venkatesh, Arjun, [arjun.venkatesh@yale.edu](mailto:arjun.venkatesh@yale.edu)



## Additional Information

1. Provide any supplemental materials, if needed, as an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be collated one file with a table of contents or bookmarks. If material pertains to a specific criterion, that should be indicated.

[Response Begins]

[Response Ends]

2. List the workgroup/panel members' names and organizations.

*Describe the members' role in measure development.*

[Response Begins]

[Response Ends]

3. Indicate the year the measure was first released.

[Response Begins]

[Response Ends]

4. Indicate the month and year of the most recent revision.

[Response Begins]

[Response Ends]

5. Indicate the frequency of review, or an update schedule, for this measure.

[Response Begins]

[Response Ends]

6. Indicate the next scheduled update or review of this measure.

[Response Begins]

[Response Ends]

7. Provide a copyright statement, if applicable. Otherwise, indicate "N/A".

[Response Begins]

[Response Ends]

8. State any disclaimers, if applicable. Otherwise, indicate "N/A".

[Response Begins]

[Response Ends]

9. Provide any additional information or comments, if applicable. Otherwise, indicate "N/A".

[Response Begins]

**[Response Ends]**